

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: NDA 20571/S9

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

MAR 29 2000

NDA Number:	20-571 suppl. S-09
Applicant:	Pharmacia & Upjohn Co.
Name of Drug:	Camptosar (irinotecan hydrochloride) Injection
Indication:	First-line therapy of metastatic colorectal cancer.
Documents Reviewed:	Vols. 1.3-1.30 dated 19 Oct 1999; data submission dated 19 Nov 1999, SE1-009 dated 10 Feb 2000
Medical Reviewer:	Isagani Chico, M.D.
Statistical Reviewer:	David Smith, Ph.D.

**APPEARS THIS WAY
ON ORIGINAL**

1. Background and Overview

In order to support labeling for the indication of treatment of first-line metastatic colorectal cancer, the sponsor submitted a supplemental NDA which is comprised of two Phase III trials and one Phase I trial. The Phase I study was submitted to support data that describe the pharmacokinetics of Camptosar when given together with 5-FU and leucovorin. We will only consider the Phase III studies in this review, as these are the studies relevant in demonstrating the effectiveness of Camptosar.

A brief summary of the pivotal studies appears below.

Study	Type	N	Arms
V303	Randomized Ph III	54	A1 Regimen: 80 mg/m ² CPT-11 + 500 mg/m ² LV + 2300 mg/m ² 5FU
		145	A2 Regimen: 80 mg/m ² CPT-11 + 200 mg/m ² LV + 400 mg/m ² 5FU bolus + 600 mg/m ² 5FU over 22 hrs
		199	A1+A2: CPT-11* / 5-FU** / LV***
		43	B1 Regimen: 500 mg/m ² LV + 2300 mg/m ² 5FU
0038	Randomized Ph III	143	B2 Regimen: 200 mg/m ² LV + 400 mg/m ² 5FU bolus + 600 mg/m ² 5FU over 22 hrs
		186	B1+B2: 5-FU / LV
		226	A (42-day course) 125 mg/m ² CPT-11
		231	B (42-day course) 125 mg/m ² CPT-11 + 20 mg/m ² LV + 500 mg/m ² 5FU
		226	C (28-day course) 5-FU / LV

*CPT-11 = Camptosar; **5-FU = 5-fluoruracil; ***LV = leucovorin.

The next section includes relevant statistical issues for these studies. The following sections will discuss these studies in more detail and will follow the following format:

1. General description of study
2. Efficacy endpoints and results
3. Summary and conclusions

The last two sections will include a summary of efficacy, overall conclusions and recommendations for the submission.

2. Statistical Issues

- An imbalance in randomization occurred in V303; a statistically significantly larger number of patients with rectal cancer and a larger number of females were randomized to Arm A. More patients with prior surgery were enrolled in Arm B (Table 3.1).
- An imbalance in randomization also occurred in V0038; a statistically significantly larger number of females were randomized to Arm C compared to the other two arms (Table 3.5).
- In V303, one may argue that it is not appropriate to perform a statistical analysis on arms that have each been combined from two separate regimens of drug (e.g., A1 vs. A2, B1 vs. B2). Even though the arms are demographically similar, there is an inherent heterogeneity among the data that may not be addressed by randomization. There are many precedents of combining regimens to construct arms for later comparison, although one should consider this element of heterogeneity when interpreting the results of such comparisons.
- With respect to comparisons of duration of response, there is an inherent selection bias. There is little evidence that the subset of responders can be predicted at baseline prior to treatment. In addition, the scheme for predicting the responding subsets will differ from treatment to treatment. Therefore, comparisons of durations of response are spurious in that they compare heterogeneous data sets; duration of response is of little use in determining drug efficacy.

- Although the Kaplan-Meier curves for survival in V303 suggest that the proportional hazards assumption may not be met, formal tests indicate that there is no evidence to reject a null hypothesis of proportional hazards when performing Cox regression analyses.
- In the QOL analyses, there was a subgroup of patients who dropped out with a consent withdrawal. As both studies progressed, this subgroup may be a source of informative missing data. However, it is very difficult to construct a reasonable model for this dropout mechanism. The sponsor did not adjust for this source of informative dropout in their analyses but this is merely one definable source among many potential sources of informative dropout in analyzing QOL data in a general oncology context.
- The quality of life analysis consisted of comparing subscale measurements at each follow up to baseline measurements using MANOVA or ANOVA techniques. However, conclusions based on these methods may be biased, due to the presence of missing data. The sponsor did perform a longitudinal quality of life analysis, which is more appropriate in determining trends over time and variance inflation due to patient dropout. However, results of these longitudinal analyses were inconclusive, favoring neither Camptosar or its control in V303 or 0038. There was also no attempt to adjust Type I error for the multiplicity of subscales considered.

3. Pivotal Phase III Trials

3.1 Description of Study V303

Study Objective: To evaluate response rate of CPT-11+5FU/LV versus 5FU/LV in patients with metastatic colorectal cancer.

Study Enrollment Period: May 1997 - February 1998

Study Design: Open label, multi-center randomized Phase III study. The only stratification factor was center.

Sample Size: Out of 385 patients randomized, 54 were randomized to regimen A1, 145 were randomized to regimen A2, 43 were randomized to regimen B1, and 143 were randomized to B2. In sum, 199 patients received a regimen containing CPT-11 and 186 patients received no CPT-11.

The sponsor assumed that the response rate for both of the 5FU/LV only arms was 35% and 50% for both of the CPT-11-containing arms. Under a two-sided χ^2 test with a Type I error of 0.05 and power of 0.80, 338 patients would detect a significant difference in response rates.

Interim Analysis: No interim analysis was specified.

Dosing:

A1 Regimen: 80 mg/m ² CPT-11 + 500 mg/m ² LV + 2300 mg/m ² 5FU
A2 Regimen: 80 mg/m ² CPT-11 + 200 mg/m ² LV + 400 mg/m ² 5FU bolus + 600 mg/m ² 5FU over 22 hrs
B1 Regimen: 500 mg/m ² LV + 2300 mg/m ² 5FU
B2 Regimen: 200 mg/m ² LV + 400 mg/m ² 5FU bolus + 600 mg/m ² 5FU over 22 hrs

See the FDA medical review of CPT-11 for further details.

Criteria for Evaluation: The primary efficacy endpoint was response rate on the intent-to-treat population. Covariates include demographic variables, weight, performance status, tumor characteristics and prior chemotherapy.

Demographic Differences at Baseline

Between arms A and B, the patients were demographically similar (i.e., no significant statistical differences between A and B) except for the following exceptions in Table 3.1.

Table 3.1. Tests for balance of randomization among the baseline demographic endpoints of Study V303.

	A1+A2	B1+B2	p-val
Gender	33.3% female	47.1% female	0.0006
Age	60.0 years, mean	57.9 years, mean	NS
WHO PS	93.4% either 0 or 1	92.5% either 0 or 1	NS
Weight loss	3.7% lost \geq 5% body wt	3.4% lost \geq 5% body wt	NS
Rectal tumors	45.5% rect/rect sigmoid	35.3% rect/rect sigmoid	0.042
Num of organs involved	85.3% either 1 or 2	90.9% either 1 or 2	NS
Sites of disease	76.8% liver	79.7% liver	NS
	26.3% lung	23.0% lung	NS
Prior adj chemo	25.8% yes	23.5% yes	NS
Prior surgery	88.9% yes	94.7% yes	0.041
Prior radiotherapy	20.2% yes	15.5% yes	NS

For the most part, patients were well-balanced at baseline.

Efficacy Endpoints

Table 3.2 shows a summary of the efficacy endpoints in study V303.

Table 3.2. Analyses of efficacy endpoints for the intent-to-treat population of Study V303.

Primary Endpoint	Arm	PR% / CR%	p-value
Response Rate	A1+A2 (CPT-11)	31.8 / 3.0	0.005
	B1+B2 (no CPT-11)	21.9 / 0.0	
Secondary Endpoints	Arm	Months (Range)	p-value
Overall Survival	A1+A2 (CPT-11)	17.4 (15.2 - 20.1)	0.032
	B1+B2 (no CPT-11)	14.1 (12.6 - 17.4)	
Time to Progression	A1+A2 (CPT-11)	4.4 (0.0+ - 11.8)	<0.001
	B1+B2 (no CPT-11)	6.2 (0.0+ - 11.8)	
Time to Treatment Failure	A1+A2 (CPT-11)	5.3 (0.4 - 15.7)	0.001
	B1+B2 (no CPT-11)	3.8 (0.4 - 11.5+)	
Time to Performance Status Deterioration	A1+A2 (CPT-11)	11.2 (0.1+ - 15.7+)	0.046
	B1+B2 (no CPT-11)	9.9 (0.0+ - 13.6+)	
Duration of Response*	A1+A2 (CPT-11)	9.3 (2.8+ - 13.1)	0.08
	B1+B2 (no CPT-11)	8.8 (3.7+ - 11.8+)	

*See note in Section 2.

Note that all significant p-values in Table 3.2 indicate a CPT-11 advantage over no CPT-11.

Duration of response is not useful in determining efficacy in that selecting the responding subgroup for both treatments is very improbable. The comparisons of response durations is similar to comparing heterogeneous data sets; very little information may be gained from an endpoint with an inherent selection bias.

Logistic Regression Analysis of Response Rate

The protocol specified the following prognostic factors that were clinically relevant: sex, WHO performance status, weight loss, number of organs involved, primary tumor sites, site of disease and time between first diagnosis to first metastases. A stepwise regression on these prognostic factors was performed with response as the dependent variable; the criteria to enter the model was a p-value less than 0.08 and the p-value required to remain in the model was 0.10. The two variables that were chosen from the stepwise procedure were weight loss and time between first diagnosis to first metastases. When treatment group (as defined by A1+A2 vs. B1+B2) was added, the three independent variables (treatment, weight loss, and time between first diagnosis to first metastases) were all significant at 0.05. The treatment arm showed that those patients with CPT-11-containing regimens had a greater response rate than those patients who received no CPT-11.

Cox Regression Analysis of Time to Progression

As a secondary analysis, a Cox model adjusted for a group of prognostic factors was fitted for time to progression. In addition to the baseline demographic factors, factors found in a previous time to progression Cox analysis were included: liver involvement, lymph node involvement, both liver and lymph nodes involved, time from first diagnosis to time until first infusion. Following the same regression analysis scheme as in the Cox regression above, a univariate analysis was performed to select the significant variables for a multiple Cox regression. Variables were selected for the multiple regression if they were statistically significant in a univariate model using an alpha of 0.10. A stepwise regression on these prognostic factors was performed with time to progression as the dependent variable; the criteria to enter the model was a p-value less than 0.08 and the p-value required to remain in the model was 0.10. After the stepwise regression, the variables remaining were age and liver involvement. When treatment group (A1+B1 vs. A2+B2) was added, treatment was significant ($p < 0.001$), age was not significant ($p = 0.11$), and liver involvement was not significant ($p = 0.067$).

Stratified Cox Regression Analysis of Survival

In an updated submission with follow-up data to 20 Dec 2000, a stratified Cox model adjusted for baseline demographic factors was fitted for survival. The stratified Cox regression analysis was performed similarly to the ones mentioned previously. The variables in the model are those that appear in Table 3.3. Notably, Arm A showed statistically significant improvement in time to progression over Arm B.

Table 3.3. Stratified Cox regression analyses of survival for the intent-to-treat population of Study V303.

Factor at Baseline	Categorization	Hazard + 95% CI	p-value
Treatment Group	A1+B1 vs. A2+B2	0.77 (0.61 - 0.98)	0.0368
Serum LDH	\leq UNL vs. $>$ UNL	0.55 (0.42 - 0.72)	0.0001
Performance Status	0 vs. $>$ 0	0.52 (0.41 - 0.67)	0.0001
Time from Diagnosis	\geq 1 vs. $<$ 1 month	0.63 (0.49 - 0.82)	0.0005
Involved Organs	1 vs. \geq 2 sites	0.73 (0.57 - 0.94)	0.0127

In a sensitivity analysis, there was statistically significant evidence that there was an interaction between treatment and performance status. When a performance-status-by-treatment-interaction term was added into the model, treatment differences remained significant at $p = 0.0012$ for the subgroup with performance status of 0 and $p = 0.9442$ for the subgroup with performance status greater than 0.

Efficacy Endpoints within Each Regimen

A clinical issue for this application is that of comparability of regimens, namely whether it is appropriate to combine two data sets with differing regimens. If regimens are similar to one another, one would expect agreement in A1 vs. B1 and A2 vs. B2 comparisons across the endpoints under consideration. We examine now the results of the individual treatments {A1,A2,B1,B2}. Note that due to the unequal distribution between regimens 1 and 2, only the number of patients in regimen 2 are reliable with respect to efficacy.

Table 3.4 shows a summary of the efficacy endpoints in study V303.

Table 3.4. Analysis of response rate and time to event endpoints for each of the four regimens for the intent-to-treat population of Study V303.

Primary Endpoint	Arm	PR% / CR%	p-value	Arm	PR% / CR%	p-value
Response Rate	A1	35.8/31.8	0.127	A2	30.3/31.1	0.021
	B1	25.0/0.0		B2	21.0/0.0	
Secondary Endpoints		Months (Range)	p-value		Months (Range)	p-value
Overall Survival	A1	19.2 (1.4 - 19.2)	0.476	A2	15.6 (0.4 - 19.9+)	0.041
	B1	14.1 (1.0 - 17.2+)		B2	13.0 (0.5 - 19.4)	
Time to Progression	A1	7.2 (0.0+ - 13.8)	0.184	A2	6.5 (0.0+ - 13.2)	0.001
	B1	6.5 (0.0+ - 12.3+)		B2	3.7 (0.0+ - 13.1+)	
Time to Treatment Failure	A1	5.4 (0.8 - 15.8+)	0.346	A2	5.1 (0.4 - 17.1+)	<0.001
	B1	5.0 (0.4 - 10.4+)		B2	3.0 (0.5 - 11.7+)	
Duration of Response*	A1	8.9 (2.8+ - 12.4)	0.043	A2	9.3 (3.9+ - 13.1)	0.880
	B1	6.7 (3.7+ - 10.3)		B2	9.5 (3.9+ - 13.1)	

*See note in Section 2.

Note that the χ^2 test was specified in the protocol as the formal method to analyze response rate. Fisher's exact test is a secondary analysis that one may consider for the small number of patients in arms A1 and B1. The p-value for Fisher's exact test for comparing responses between arms A1 and B1 is $p = 0.137$. The p-value for Fisher's exact test for comparing responses between arms A2 and B2 is $p = 0.021$.

The difference in response rate between A1 and B1 is similar to that between A2 and B2. The large p-value ($p = 0.127$) for comparing responses between Arms A1 and B1 may be due to a relatively smaller sample size as A2 and B2. Similarly, the difference in median survival observed between Arm A1 and B1 is approximately five months, which is greater than that observed between Arms A2 and B2 (approx. 2.6 months).

Quality of Life

To assess quality of life (QOL), the sponsor administered the EORTC QLQ-C30 instrument in both treatment arms. In the following discussion, the unit of time that we consider for V303 will be cycles of treatment. Figure 3.1 shows the proportion of compliance to the QOL instrument for V303.

Figure 3.1. Proportion of compliance to the QOL instrument in Study V303.

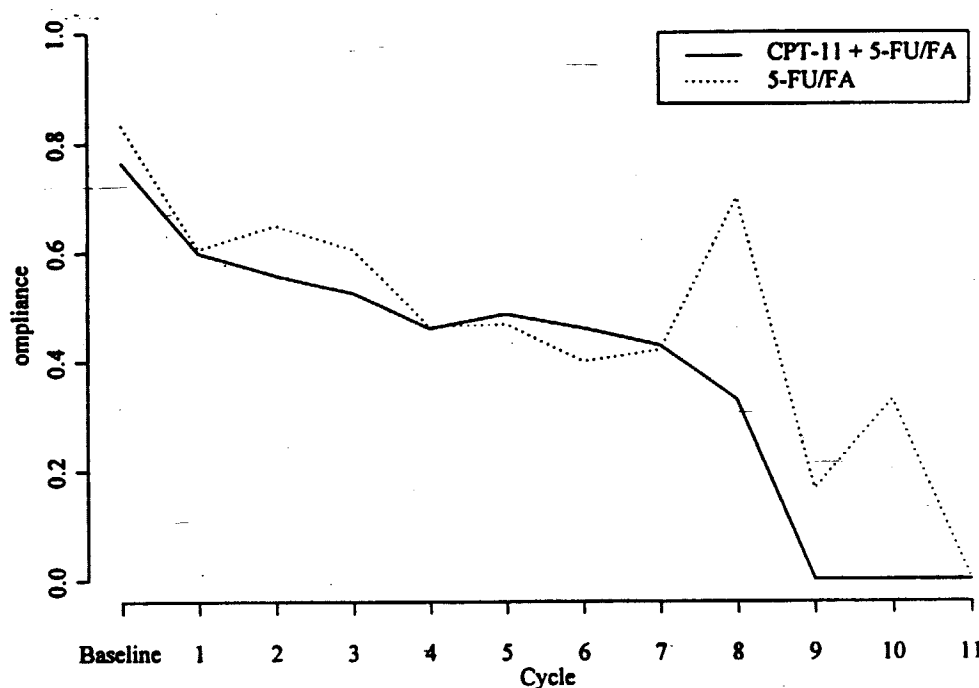


Figure 3.1 indicates that compliance to the QOL instrument was similar between A and B treatments through approximately Cycle 7. Very few patients remained on the study subsequent to Cycle 7.

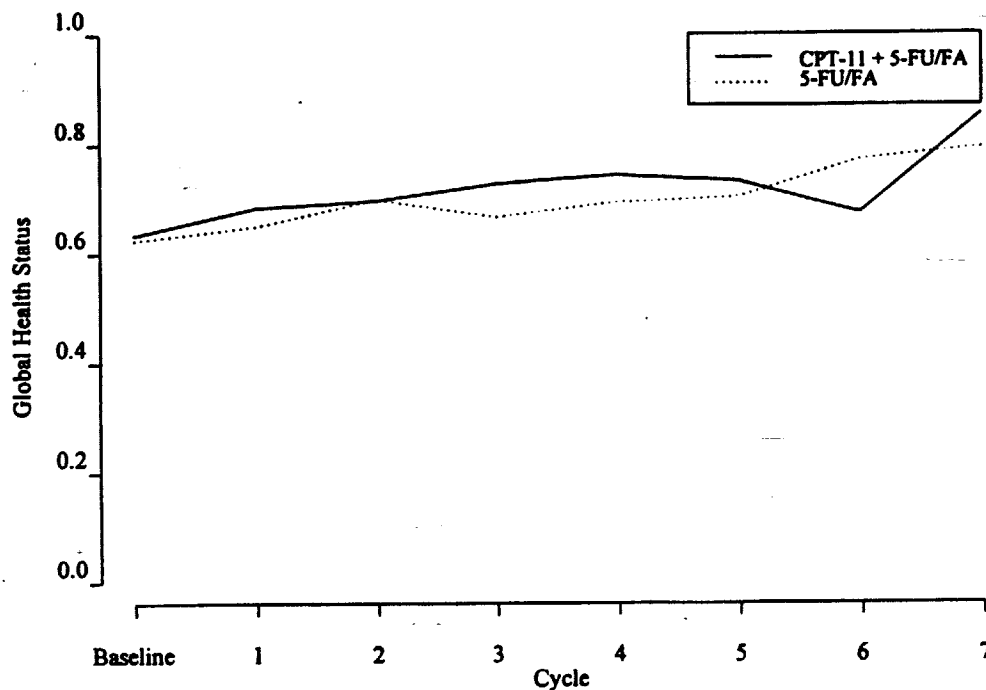
At baseline, the QOL scales showed no statistically significant differences between the treatment means. Therefore, no adjustments to correct for baseline differences was necessary.

The primary analysis set that was defined for this study were those patients who had an evaluable QOL score above 10% of the initial number of patients in each arm. A formal analysis of the comparability of the missingness of the data was subsequently performed. Thirty-one patients who only filled out a questionnaire at baseline were dropped from the analysis. The sponsor examined the relationship between patient dropout and the primary reason for dropout within the same cycle. Progressive disease was the primary reason for dropout; progression occurred more often in arm B than arm A. However, there were relatively more patients who dropped out for adverse events in arm A than arm B. The sponsor concluded that the dropout mechanism was not random among arm A and B, and there was an imbalance of dropout such that comparisons of QOL between the two arms would be biased.

The sponsor proposed several methods to re-examine the QOL data in the presence of a bias from dropout. Two such methods relied on differing interpretations on how to adjust for dropouts due to progressive disease. The first of the two methods used Last Observation Carried Forward (LOCF) as an imputation scheme; the second of the two methods imputed the mean of the worst scores of progressive patients (zero if death).

Global Health Status on the EORTC QLQ-C30 is a measurement of overall QOL. Figure 3.2 shows the mean scores of patients' Global Health Status scores from baseline to Cycle 7. Note that these scores are highly biased due to the excessive informative dropout among the patients.

Figure 3.2. The results of the Global Health Status raw scores (*i.e.*, no imputation) through Cycle 7 for V303. An increase in Global Health Status implies an increase in patient benefit.



The following two figures correspond to the sponsor's imputation analyses. Figure 3.3 shows the means of patients' Global Health Status scores when missing data is imputed using a last observation carried forward (LOCF) scheme. Figure 3.4 shows the means of patients' Global Health Status scores when missing data is imputed using each patient's worst score. These data reflect only those measurements taken through Cycle 7. It is notable that Figure 3.3 and 3.4 appear to be similar, which implies that there is a correlation between the last observation being carried forward and each patient's worst score.

APPEARS THIS WAY
ON ORIGINAL

Figure 3.3. The results of the Global Health Status scores with LOCF values imputed for missing values through Cycle 7 for V303. An increase in Global Health Status implies an increase in patient benefit.

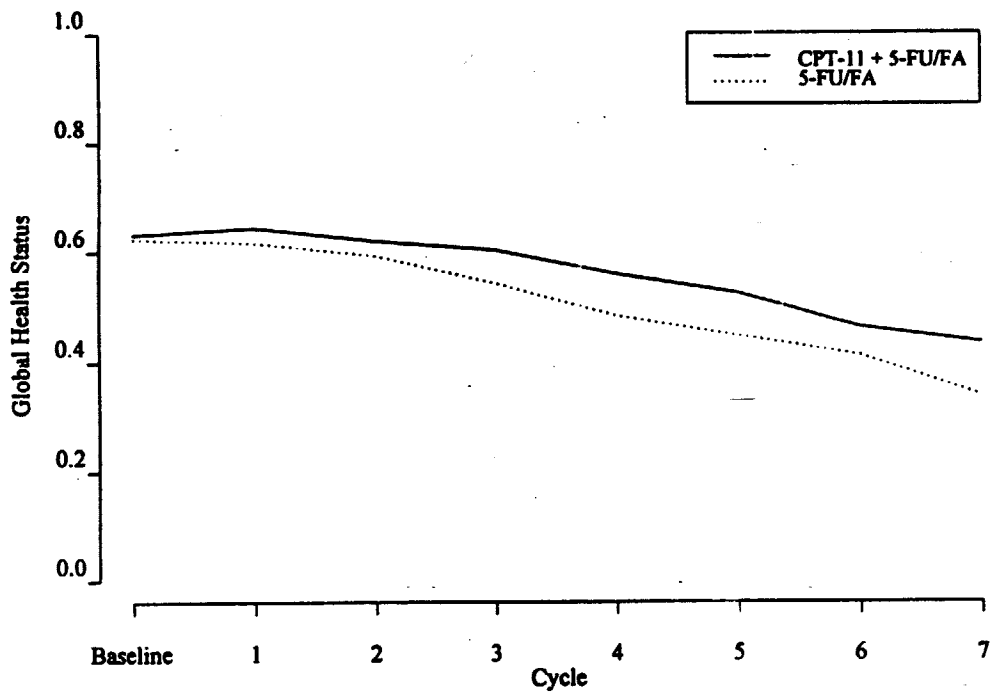
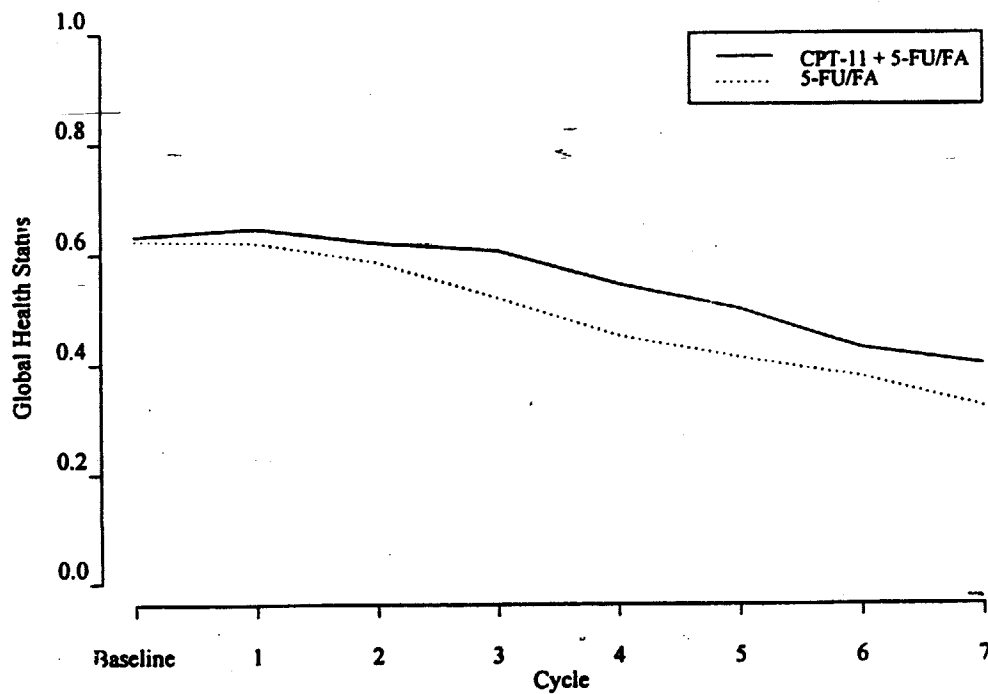


Figure 3.4. The results of the Global Health Status scores with LOCF values imputed for missing values through Cycle 7 for V303. An increase in Global Health Status implies an increase in patient benefit.



In Figures 3.3 and 3.4, the difference ($\text{Arm A}_{\text{QOL}} - \text{Arm B}_{\text{QOL}}$) is positive across all time intervals and is time-dependent. One conclusion that we may draw is that the overall bias induced from progressive disease and adverse event missing data causes a dropout bias against Arm A when it is compared with Arm B. Further formal analyses showed that differing QOL trends emerge in Arm A when one considers QOL as a function of Cycle. In Arm B, however, there is little dependence on Cycle when examining QOL trends.

In a more extensive treatment of modeling patient dropout with respect to QOL, the sponsor performed a mixed models analysis using time period as a repeated measures factor. Generally, there were trends in favor of Arm A, although overall, the trends were not significant.

The sponsor's comprehensive examination of QOL in V303 resulted in evidence that raises several concerns with respect to interpreting QOL comparisons between Arm A and B. Namely, there is evidence that dropout confounds any conclusions one can draw from formal tests between the QOL of A versus the QOL of B. The sponsor and this reviewer agree that any results with respect to QOL comparisons must be interpreted with caution, as they are likely to be biased.

3.2 Description of Study 0038

Study Objective: To evaluate time to progression of CPT-11+5FU/LV versus 5FU/LV in patients with metastatic colorectal cancer.

Study Enrollment Period: May 1996 - May 1998

Study Design: Open label, parallel-group, multi-center randomized Phase III study. The stratification factors were age, prior adjuvant 5-FU based therapy, time from initial diagnosis, and performance status.

Sample Size: Out of 683 patients randomized, 226 were randomized to regimen A, 231 were randomized to regimen B, and 226 were randomized to regimen C.

The sponsor assumed that the median time to progression would be 5 months for the 5FU/LV arm (Arm C). A 40% increase in median time to progression would translate to a 7 month median time to progression for the CPT-11/ 5FU/LV arm (Arm B). Therefore, 162 events per treatment arm are required to detect this difference under 85% power and a two-tailed Type I Error of 0.05. After adjusting for non-compliance and the timing of the primary analysis, the sponsor adjusted this sample size estimate upwards, resulting in a sample size of 220 patients per arm.

Interim Analysis: No interim analysis was specified.

Dosing:

A Regimen (42-day course): 125 mg/m ² CPT-11 weekly
B Regimen (42-day course): 125 mg/m ² CPT-11 weekly + 20 mg/m ² LV weekly + 500 mg/m ² 5FU weekly
C Regimen (28-day course): 20 mg/m ² LV daily x 5 + 425 mg/m ² 5FU weekly

See the FDA medical review of CPT-11 for further details.

Criteria for Evaluation: The primary efficacy endpoint was time to progression in the intent-to-treat population. The primary comparison was specified to be between Arm B and Arm C. In addition to the specified stratification factors, covariates include gender, ethnic origin, and site of primary tumor.

Demographic Differences at Baseline

Table 3.5 shows the comparisons among demographic values at baseline.

Table 3.5. Comparisons among the three arms of study 0038 across selected demographic variables.

	A	B	C
Gender	35.4% Female	34.2% Female	44.7% Female
Age	61 yrs (median)	62 yrs (median)	61 yrs (median)
ECOG PS	91.6% either 0 or 1	84% either 0 or 1	86.3% either 0 or 1
Rectal tumors	14.6% rectal	16.5% rectal	13.7% rectal
Num of organs involved	90.2% either 1 or 2	89.1% either 1 or 2	88.9% either 1 or 2
Liver involvement	83.2% involved liver	81.8% involved liver	81.9% involved liver
Prior adj chemo	10.2% yes	10.8% yes	8.0% yes
Prior surgery	90.2% yes	91.0% yes	90.2% yes
Prior radiotherapy	1.3% yes	3.0% yes	2.2% yes

For the most part, patients were well-balanced at baseline except for the finding that there were statistically significantly more women enrolled in Arm C than Arm B.

Efficacy Endpoints

Table 3.6 shows a summary of the efficacy endpoints in study 0038.

Table 3.6. Analyses of efficacy endpoints for the intent-to-treat population of Study 0038. Note that Arm A was not used in formal comparisons although we report the descriptive statistics here for completeness. The p-value column refers to the comparison between Arm B (CPT-11) and Arm C (no CPT-11).

Primary Endpoint	Arm	Median in months & 95% CI	p-value
Time to Progression	B	5.4 (3.4 - 8.0)	0.004
	C	3.9 (3.7 - 4.6)	
	A	4.2 (3.9 - 5.0)	---
Secondary Endpoints	Arm	Med. & 95% CI	p-value
Overall Survival	B	14.8 (12.3 - 17.1)	0.042
	C	14.0 (11.1 - 14.6)	
	A	12.0 (11.3 - 13.5)	---
Time to Treatment Failure	B	5.4 (4.3 - 6.8)	0.001
	C	3.7 (3.0 - 4.3)	
	A	3.2 (2.9 - 4.1)	---
		CR&PR, 95% CI	
Confirmed Response Rate	B	39.4 (33.0 - 46.0)	<0.0001
	C	20.8 (15.7 - 26.7)	
	A	18.1 (13.3 - 23.8)	---

Note that both primary and secondary endpoints show a statistically significant advantage in favor of Arm B over Arm C. The point estimates for Arm A are generally slightly inferior than those in Arm C, although this was not tested formally.

Cox Regression Analysis of Time to Progression

As a secondary analysis, a Cox model adjusted for baseline demographic factors was fitted for time to progression. A univariate analysis was performed to select the significant prognostic variables for a multiple Cox regression. Variables were selected for the multiple regression if they were statistically significant in a univariate model using an alpha of 0.10. After the forward selection, the variables remaining were those that appear in Table 3.7. Notably, Arm B showed statistically significant improvement in time to progression over Arm C.

Table 3.7. Cox regression analyses of time to progression for the intent-to-treat population of Study 0038.

Factor at Baseline	Categorization	Hazard + 95% CI	p-value
Treatment	Arm B vs. Arm C	0.64 (0.51 - 0.79)	0.0001
Serum LDH	≤ UNL vs. > UNL	0.60 (0.47 - 0.76)	0.0001
Num. Involved Organs	1 vs. ≥ 2 sites	0.63 (0.50 - 0.80)	0.0001
Performance Stat.	0 vs. > 0	0.74 (0.59 - 0.93)	0.0088
Serum Tot. Bilirubin	≤ UNL vs. > UNL	0.56 (0.35 - 0.89)	0.0132
Hemoglobin	< 11 vs. ≥ 11 g/dL	0.74 (0.58 - 0.95)	0.0157
Age	< 65 vs. ≥ 65 years	0.78 (0.63 - 0.98)	0.0315

In a subsequent analysis, there was statistically significant evidence that there was an interaction between treatment and serum LDH. When a serum-LDH-by-treatment-interaction term was added into the model, treatment differences remained significant at $p < 0.001$.

Multiple Regression Analysis of Confirmed Objective Response

A forward-selection regression analysis was performed to select the significant variables for a multiple regression. Variables were selected for the multiple regression if they were statistically significant in a univariate model using an alpha of 0.10. After the forward selection, the variables remaining were those that appear in Table 3.8. Notably, Arm B showed statistically significant improvement in time to progression over Arm C.

Table 3.8. Multiple regression analyses of confirmed objective response for the intent-to-treat population of Study 0038.

Factor at Baseline	Categorization	Hazard + 95% CI	p-value
Treatment	Arm B vs. Arm C	2.99 (1.91 - 4.67)	0.0001
Time from Diagnosis	< 1 vs. ≥ 1 month	0.61 (0.38 - 0.98)	0.0339
Num. Involved Organs	1 vs. ≥ 2 sites	1.70 (1.06 - 2.72)	0.0267
Performance Stat.	0 vs. > 0	2.08 (1.35 - 3.20)	0.0008
Serum Tot. Bilirubin	≤ UNL vs. > UNL	6.08 (1.35 - 27.4)	0.0132

There were no statistically significant interactions between treatment and the other prognostic factors given in Table 3.8.

Stratified Cox Analysis of Survival

In an updated submission with follow-up data to 20 Dec 2000, a stratified Cox model adjusted for baseline demographic factors was fitted for survival. The regression analysis was performed similarly to the ones mentioned previously. The variables in the model are those that appear in Table 3.9. Notably, Arm B showed statistically significant improvement in survival over Arm C.

Table 3.9. Stratified Cox regression analyses of survival for the intent-to-treat population of Study 0038.

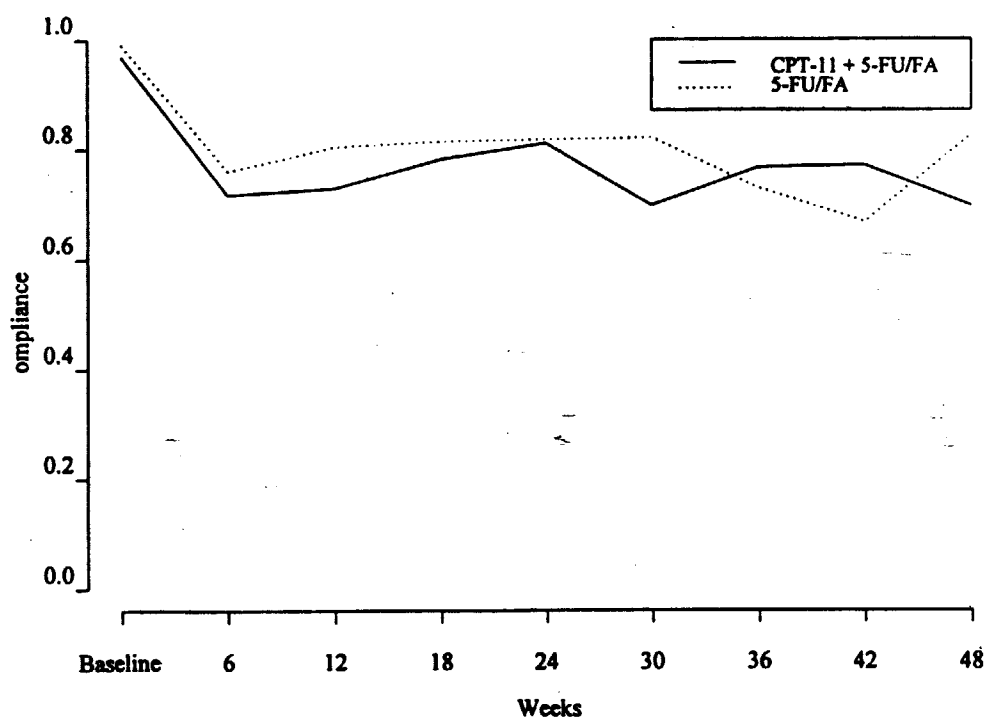
Factor at Baseline	Categorization	Hazard + 95% CI	p-value
Arm B vs. Arm C		0.80 (0.66 - 0.99)	0.0372
Serum LDH	≤ UNL vs. > UNL	0.49 (0.38 - 0.62)	0.0001
Serum Tot. Bilirubin	≤ UNL vs. > UNL	0.58 (0.37 - 0.91)	0.0185
WBC per mm ³	< 8x10 ³ vs. ≥ 8x10 ³	0.66 (0.53 - 0.82)	0.0002

There were no statistically significant interactions between treatment and the other prognostic factors given in Table 3.9.

Quality of Life

To assess quality of life (QOL), the sponsor administered the EORTC QLQ-C30 instrument in both treatment arms. In the following discussion, the unit of time that we consider for Study 0038 will be weeks of treatment. Figure 3.5 shows the proportion of compliance to the QOL instrument for Study 0038.

Figure 3.5. Proportion of compliance to the QOL instrument in Study 0038.



As previously described for the QOL analysis for V303 in Section 3.1, the sponsor performed two analyses other than an analysis of the raw QOL data. In a repeated measures ANOVA, there were no statistically significant differences between arms B and C across all QOL endpoints using Zwinderman's method of imputation for missing values. When worst changes from baseline were analyzed, a statistically significant difference was found between pain and role functioning, each in favor of arm B. This is a somewhat conservative analysis, but it is difficult to extrapolate this result into compelling patient benefit.

The sponsor's comprehensive examination of QOL in 0038, like V303, raises several concerns with respect to interpreting QOL comparisons between Arm B and C. Namely, there is evidence that dropout confounds any conclusions one can draw from formal tests between the QOL of B versus the QOL of C. The sponsor and this reviewer agree that any results with respect to QOL comparisons must be interpreted with caution, as they are likely to be biased.

4. Summary and Conclusions

These studies were designed to test the superiority in the primary endpoint of response rate of CPT-11 versus control arms of 5FU/LV regimens in study V303 and superiority of time to progression of CPT-11 versus control arms of 5FU/LV in study 0038. In the intent-to-treat population of the V303 and 0038 Phase III trial, CPT-11 was statistically significantly better across major efficacy endpoints, such as response rate, survival, time to progression and time until performance status deterioration (see Tables 3.2 and 3.6).

Additionally, Cox modeling was performed both by the sponsor and by this reviewer.* For the Cox analyses on survival and time to progression, treatment arm was statistically significant in both studies.

Patients received the EORTC QLQ-C30 questionnaire to measure QOL in both studies. In each studies' analyses of QOL, there was evidence that there was non-ignorable dropout and that any decisions regarding QOL are likely to be biased. Formally, the QOL results for both studies are inconclusive, but descriptively speaking, there seems to be a general trend favoring QOL improvement or less-worsening in favor of the CPT-11 arm.

5. Overall Recommendations and Conclusions

In the two Phase III trials included in this submission, CPT-11 was statistically significantly superior to 5FU/LV in both primary endpoints (response rate in V303 and time to progression in 0038). There is substantial evidence to conclude that CPT-11 is statistically superior to 5FU/LV across nearly all major endpoints, including survival.

Because of bias due to dropout, one may not formally conclude with respect to QOL in both studies. Both the sponsor's and the FDA's analyses showed the evidence of non-ignorable dropout. Although the sponsor performed an exemplary analysis of QOL, there is no compelling statistical evidence that there are differences between CPT-11 and 5FU/LV with respect to QOL improvement.

It is this reviewer's opinion that CPT-11 has demonstrated efficacy for the proposed indication based on the two trials that were submitted.

JS/

David Smith, Ph.D.
Mathematical Statistician

Concur:

Dr. Chen

/S/

3/28/00

Dr. Mahjoob

~~/S/~~

3/28/2000

cc: Archival NDA 20-571
HFD-150 / Ms. B. Atkins, Project Manager
HFD-150 / Dr. I. Chico
HFD-150 / Dr. G. Williams
HFD-344 / Dr. B. Barton
HFD-710 / Dr. G. Chi
HFD-710 / Dr. K Mahjoob
HFD-710 / Dr. G. Chen
HFD-710 / Chron

Smith / 28 March 2000 / Word / c:\IND&NDA review\20571cr_final.doc
This review consists of 15 pages of text.